

Virtual Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises

The increasing availability of software for creating real-time simulations of musical instrument sounds allows for the design of new visual and sounding media. These past decades have especially focused on the control of real and virtual instruments by natural gestures. In this paper, we present and extensively evaluate a framework (Figure 1) for the control of virtual percussion instruments, by modeling and simulating virtual percussionists gestures. By positioning the virtual performer at the center of the gesture-sound synthesis system, we aim at providing original tools to analyze and synthesize instrumental gesture performances. Our physics-based approach for gesture simulation brings some insight into the effect of biomechanical parameters of the gesture on the instrumental performance. Simulating both gesture and sound by physical models leads also to a coherent and human-centered interaction and provides new ways of exploring the mapping between gesture and sound. The use of motion capture data enables the realistic synthesis of both pre-recorded and novel percussion sequences from the specification of gesture scores. Such scores involve motion editing techniques applied to simple beat attacks. We therefore propose an original gesture language based on the instrumental playing techniques. This language is characterized by expressivity, interactivity with the user, and the possibility to take into account co-articulation between gesture units. Finally, providing 3D visual rendering synchronized with sound rendering allows us to observe virtual performances to the light of real ones, and to qualitatively evaluate both pedagogical and compositional capabilities of such a system.

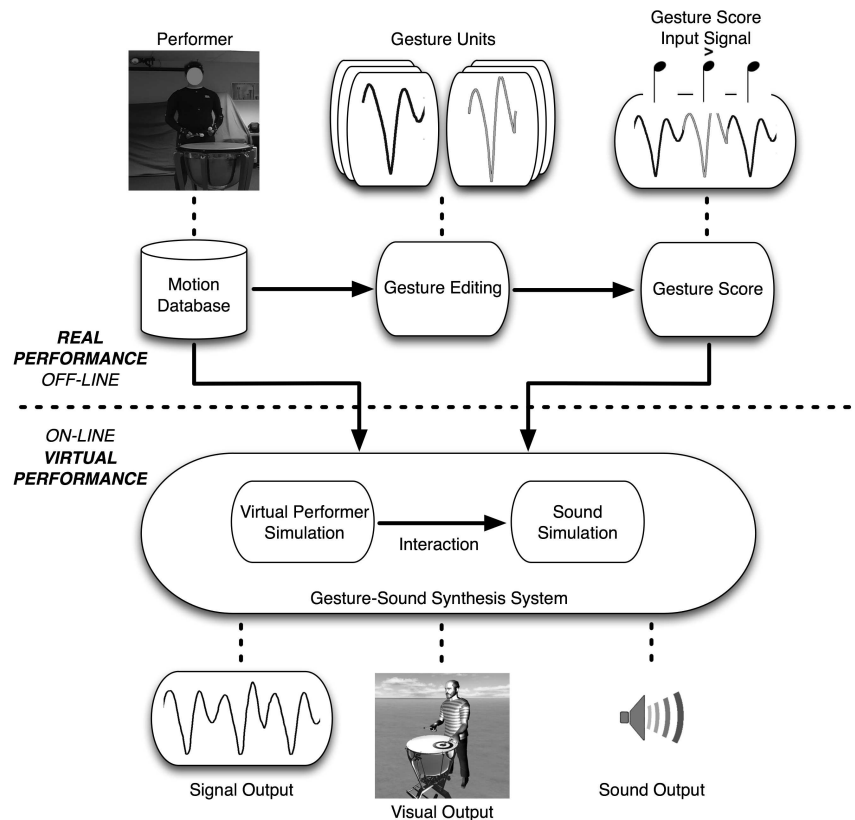


Figure 1: Global framework, from the off-line editing of a gesture score with its corresponding input signal, to the real-time simulation with visual and sound feedback.

Background

Digital musical instruments have been widely studied during the past decades, focusing mostly on the elaboration of new interfaces for controlling sound processes (Miranda and Wanderley 2006). The design of these new musical instruments relies fundamentally on the input gesture they can track and both sensor and camera-based motion capture systems have become widespread solutions for tracking instrumental gestures (Kapur et al 2003). More specifically about percussion tracking, one can find comprehensive comparisons of tracking solutions (Tindale et al 2005, Benning et al 2007). Such systems may also be used for the analysis of performer gestures (Dahl 2004, Bou  nard et al 2010a), where a good understanding of movements leads to the identification of gesture parameters that may be used for

interacting with sound synthesis processes.

Nevertheless, motion capture data intrinsically present several drawbacks. The recorded motion is dependent on the uniqueness of the interaction situation under study, in the sense that it is difficult to extrapolate it to new instrumental situations. With such data-based approaches, it is indeed far from straightforward to go beyond the recorded data and reuse it to synthesize adaptive and realistic new performances. Moreover, these systems do retrieve kinematic motion data but fall short in retrieving the physical aspects of the recorded situation. The interaction between such kinematic data with sound relies then on non-intuitive multi-dimensional correspondences (Dobrian and Koppelman 2006). Therefore, a promising research trend consists in providing motion models that could interact with sound synthesis processes.

Developments in sound synthesis have given rise to various methods to generate percussive sounds. Specifically, physics-based synthesis of percussive sounds has involved the modeling of a hammer, collision and sliding excitations (Avanzini and Rochesso 2004, Avanzini 2007), as well as drum skins (Chuchacz et al 2007). However, their main limitation seems to lie in the way they are controlled. Despite a few early attempts to approach this issue, it is still not clear how to formally relate these models to the excitation by a (real or virtual) performer.

Most of works have involved the interaction of a virtual percussion with a real performer (Mäki-Patola 2005), and contributions exploring the modeling of the equivalent gestures by the definition of a synthetic performer are fairly new. These are either based on real-world robotic models or virtual characters.

Robotic models have been used for synthesizing percussion performances (Weinberg and Driscoll 2006, Hoffman and Weinberg 2010, Kapur et al 2010). Such models generally include the design of mechanical actuators driven by various control solutions (electro-valve, solenoids, motors) for putting robots into motion,

while multi-agents approaches can be involved to specify or make emerge collaborative rules between robots (Eigenfeldt and Kapur 2008).

More related to our work among virtual models are the contributions involving a slowly evolving mechanical model (Gibet and Florens 1988), as well as the simulation of a simplified articulated arm hitting a vibrating membrane (Gibet and Marteau 1990). More recent attempts to overcome this limitation involve the animation of virtual models from MIDI data flows (Hänninen et al 1996, Lytle 2001), or from rigid body simulations with the corresponding sound synthesis (van den Doel et al 2001, O'Brien et al 2002).

Regarding either regarding robotics-inspired or virtual solutions, their main limitation seems to be related to the difficulty to compare to motion data extracted from real performances. Furthermore these do not address a complete physics-based modeling of instrumental actions and subtleties that a virtual performer could use. Our contribution addresses these issues and proposes a physics-based model of a percussionist that can be driven by real percussion (timpani) data. It enables the synthesis of pre-recorded as well as novel percussion sequences by the specification of a gesture score, the latter consisting of a choice of pre-recorded gesture units.

The paper is organized as follows. In a first time, we present the basic techniques related to timpani performance as well as the building of a motion capture database. We then introduce the global architecture of our framework, from the editing of a gesture score and the real-time simulation of percussion gestures, to its mapping to sound synthesis processes for synthesizing whole percussion performances. A qualitative evaluation of our framework is then conducted, which consists in the synthesis of percussion sequences composed of gesture units from the database. These sequences reflect both percussion exercises that are already available in the database but also exercises that were not initially captured. We then discuss the advantages and limitations of our approach, and finally conclude with

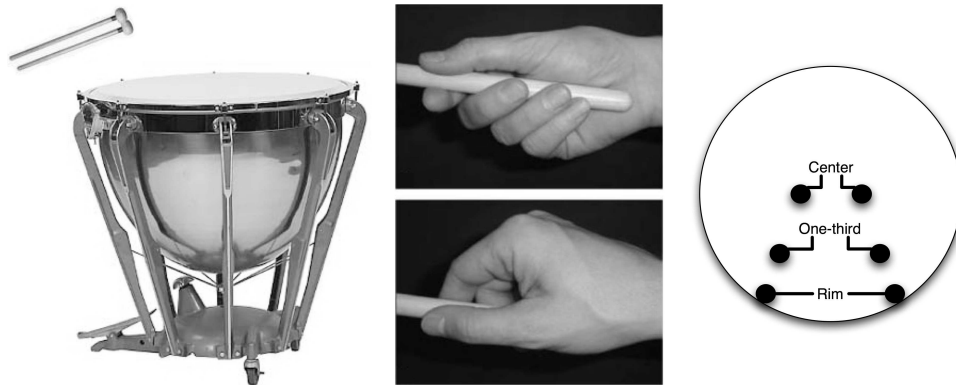


Figure 2: Timpani performance, (left) timpani instrument and mallets, (middle) French (top) and German (bottom) grips, (right) theoretical beat impact locations.

further perspectives.

Timpani Performance

In this work, we focus on a particular type of percussion instrument, timpani (Figure 2). Timpani related equipment is mainly composed of a bowl, a head and mallets. In general, timpanists have to cope with several timpani (usually four) with bowls varying in size (Frieze and Lepak 1985, Instrumentalist 1984). As for timpani mallets, they consist of a shaft and a head, and are designed in a multitude of lengths, weights, thicknesses and materials, their choice being of great importance (Instrumentalist 1984).

Timpani performers use a wide range of playing techniques. First, there are two main strategies for holding mallets: the French grip (also called "thumbs-up") and the German grip (or "matched" grip). These strategies imply different positions of the hand (vertical palm with the French grip, horizontal with the German grip), and thus different motions of the wrist and of the fingers. Moreover, the use of the limbs differ: the French grip imposes larger movement amplitudes of the arms and forearms. It should be noted however that many timpanists use techniques

combining some features of the two main techniques described here.

Performers commonly use various gesture articulation modes (legato, tenuto, accent, vertical accent and staccato), as well as three distinct locations of impacts (one-third of membrane radius, center and rim). The most used is the one-third location, producing a full sound with substantial resonance. The center (sharp and muffled attack with barely any resonance) and rim (metallic sound and a resonance bringing out mostly high frequencies) are used less often, but have been adopted by composers since Elliott Carter (Carter 1968).

Motion Capture Protocol and Database

As initially proposed in (Bouënard et al 2008a), we captured the motion of several timpani performers using a Vicon infra-red camera tracking system and a standard DV camera that allow both the retrieval of gesture and sound. Percussionists used a lycra suit fitted with markers placed according to the marker positions of Vicon's Plug-in Gait model. In order to retrieve beat impacts, markers have also been placed on the mallets. It should be noted that the placement of markers on mallets can have an impact on the recorded performance since it can slightly change their balance. The balance between left and right mallets can also be altered since different placements have been used, as a means of recognizing left and right mallets during the processing of motion data. A careful choice of the size of the timpani has been done as regards to capture conditions. A 23" timpani has therefore been chosen to minimize the occlusion of markers by the timpani bowl.

Three percussionists were asked to perform a pre-defined capture protocol consisting of a single stroke roll for each playing mode. For each mode, performers were asked to change the location of the beat impact according to the three locations described in the previous section. In total, fifteen examples of timpani exercises were performed for each percussionist (each with five beats per hand). The performers

had varied degrees of musical expertise (from Professor to Master Student or Undergraduate Student), used different grip strategies (French or German), and also differed in dominant handedness.

The choice of the sampling rate used for the acquisition of percussive gestures is important for the quality of the capture data, mostly because of the short duration of the beat impact (Wagner 2006). With high sampling rates (500Hz and above), one can expect to retrieve beat articulations more accurately, but the spatial range is significantly reduced so that it may be difficult to capture the whole body of the performer in a limited performance space (medium sized room), given a reasonable number of IR cameras (in our case, six cameras were used). For this project, a compromise was chosen by setting the cameras at 250Hz, allowing both full-body capture as well as a reasonably high sampling rate for capturing reliably mallet beat impacts.

Gesture-Sound Synthesis System

The proposed system integrates both visual and sound feedback through the real-time simulation of instrumental percussion gestures from a user-defined sequence of gestures, a "gesture score". The global architecture (Figure 1) involves three steps: a) the editing of a sequence of instrumental gestures using data from a motion capture database, b) the physics simulation of the chosen instrumental gestures, and c) interaction strategies between simulated gesture and the sound synthesis process.

Editing and Composition of Gesture Scores

The basis of the editing process is highly inspired from existing event-based works in representing music-related materials. One of the most widespread examples is the approach that was adopted in the design of the MIDI protocol. Similarly, we define a set of events that can be used in our system for synthesizing percussion

performances. However, the drastic difference is that our system uses the manipulation of elementary gesture events. In fact, these gesture events consist of motion signals coming directly from collected data presented in section "Timpani Performance", and we will therefore refer to them as gesture units. Figure 1 depicts the use of such gesture units in our synthesis system. The editing and assembling of these canonical gesture units lead to a gesture score, whose equivalent signal is then simulated by our system.

As proposed in (Ramstein 1991, Bouënard et al 2008b), such gesture units can be obtained from the segmentation of each multi-beat percussion signal into single beat-to-beat units (*i.e.* the motion between two beat impacts) by examining the physical activity of the performer-instrument interaction. However, in a simulation context, problems may arise from the assembling of such beat-to-beat gesture units. This segmentation creates two types of problems. The first one is related to the fact that linking two gesture units at the moment of the beat impact may create unwanted kinematic discontinuities, concerning the position and orientation (and their derivatives) of both mallet extremities and body segments. Another problem that can occur is the alteration of beat impact profiles. Linking two gesture units at the beat impact would lead to separate impact profiles in two phases, which may result in the modification of the action/reaction mechanical nature of the impact. Finally, placing the articulation at the exact moment of the beat impact is at odds with the usual decomposition of instrumental gestures into preparatory, interaction and retraction gestures. Such gesture units are analogous to movemes (Bregler 1997) and dynemes (Green and Guan 2004) in motion pattern recognition, as they are the source of high-level parameters retaining motion variations that can distinguish between different executions of a motion (Bouënard et al 2010a).

We therefore translate gesture units from beat-to-beat to beat-centered units. The articulation point is then placed between the retraction phase of the previous

gesture unit and the preparatory phase of the next one. Because even such a unit representation may lead to discontinuities in position and velocity between two gesture units, we ensure a C^1 continuity on drumstick extremity trajectories. Examples of beat-centered gesture units for legato and accent articulation modes are presented in Figure 1, with the corresponding simulation of the two-units assembling.

Virtual Performer Simulation

The physics modeling and control of the virtual percussionist is divided in two parts: (a) an offline process for specifying the physics model from the anthropometry of the captured percussionist, and (b) the real-time physics-based control loop achieving the simulation of the previously edited gesture score.

The physics-based model of the virtual percussionist is composed of a set of rigid solids linked by mechanical joints, considered as the virtual performer skeleton. The biomechanical properties of this skeleton (solids' length, mass, density, inertia) are initialized by motion capture data. The motion of every solid composing the physics model is simulated by solving the physical motion equations thanks to the Open Dynamics Engine library. The time step used for simulating the synthesized exercises presented in this paper is about 1 milli-second.

Our control paradigm operates directly from the mallet extremity trajectories infers joints angles (such as shoulder, elbow, wrist) from mallet trajectories. In a extracted from the gesture score, and solves two cascaded inverse problems (Bou  nard et al 2009a, 2009b, 2010b). In a first time, an Inverse Kinematics module second time, an Inverse Dynamics module exploits the joints angles previously computed for driving the mechanical joints of the virtual performer. For each mechanical joint, the physics control loop involves its angular state and target, and computes the necessary torques to be applied on the linked solids for reaching the

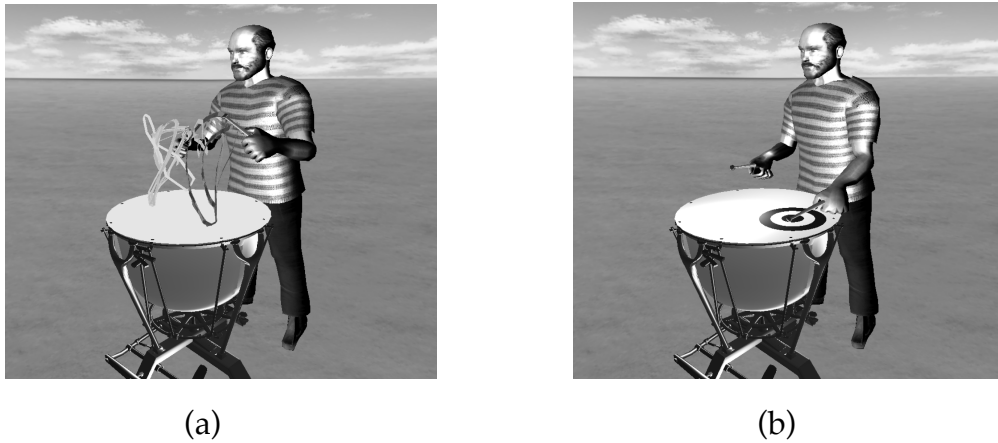


Figure 3: Visualization of (a) mallet trajectories and (b) impact position on the drum membrane

target angle. It should be noted that mechanical joints are modeled as springs with damping and stiffness coefficients that are used to model tension and relaxation factors of the virtual performer.

Multimodal Simulation and Interaction

The physics layer is exploited to express the mapping between the simulation of percussion gestures and sound synthesis methods at the physics level. The interaction between simulated percussion gestures and sound synthesis involves a collision detection algorithm that retrieves information on the impact position, velocity and force (direction and amplitude). Such interaction model can therefore be applied to various types of sound output: pre-recorded percussive sounds, signal-based (wavetable) and physics-based sound synthesis. For the physics-based sound synthesis, we experimented with a modal synthesis approach, using Modalys (Adrien 1989; Ellis et al 2005).

Two levels of interaction are involved in the simulations presented in this paper. The first one is a one-to-one mapping between the virtual performer simulation and the sound simulation. The vibrating membrane parameters are its size, stiffness, Young's modulus, and are shared between the two simulations. Beat

impact information is composed of the location on the membrane as well as the attack velocity and force. The second level of interaction is the communication of such beat impact parameters in real-time. This communication is achieved by using the Open Sound Control protocol, therefore helping the broadcast of parameters independently of the number and type of machines involved for the simulations.

As for visual outputs, users may explore the percussion performance space and visualize different features of interest (Figure 3), such as static information attached to the joints, mallets' 3D trajectories, varied camera views, virtual character rendering, as well as the visualization of impact forces and positions.

Musical Evaluation of Virtual Percussion Performances

One of the most interesting outcomes of such framework is the possibility of handling and assembling heterogeneous performances using a combination of a few percussion gesture units. Thanks to the physics simulation of the virtual performer, the issue of gesture articulation between movement units is partly addressed by the physics engine, thus leading to a more natural sequence of movements. Although, the resulting articulation is not necessarily equivalent to real performer techniques, it will nevertheless be a physically plausible solution to the problem. To evaluate the proposed framework, we have simulated an extensive set of musical examples consisting of several timpani exercises. The percussion exercises to be simulated as well as the resulting simulations were chosen and evaluated by the last author (who is a university percussion professor and active performer) by focusing both on the visual and auditory feedback, first simultaneously and then separately. The resulting exercises are available in the DVD submitted jointly to the paper. These exercises are divided in two main categories: validation and extrapolation exercises.

Throughout this section, all percussion exercises exploit gesture units for the French grip from the motion capture database described in section "Timpani











Articulations	Legato 	Tenuto 	Accent 	Vertical accent 	Staccato 
Impacts	One-third 	Center 	Rim 	Left 	Right 

Table 1: Timpani playing notation

Performance". The notation corresponding to timpani playing variations is described in Table 1, and is used as the standard notation throughout the rest of this paper for describing percussion exercise scores.

Validation Exercises

The first category, validation exercises, aims at verifying the accuracy with which the proposed framework can synthesize percussion movements (and sounds) similar to the captured movements in the database. They consist in independent exercises that explore variations on the type of grip, the type of playing mode, as well as the position of the impact on the drum membrane (Figure 4). In this section we analyze how the system manages to connect independent gesture units taken from the database, where each exercise consists on a sequence of similar types of strokes. We particularly simulated the exercises depicted in Figure 4, several sequences of (a) legato strokes, (b) tenuto strokes, (c) accent strokes, (d) vertical accent strokes, and (e) staccato strokes, all at the one-third location, and several sequences of legato strokes at (f) the one-third, (g) center, and (h) rim locations.

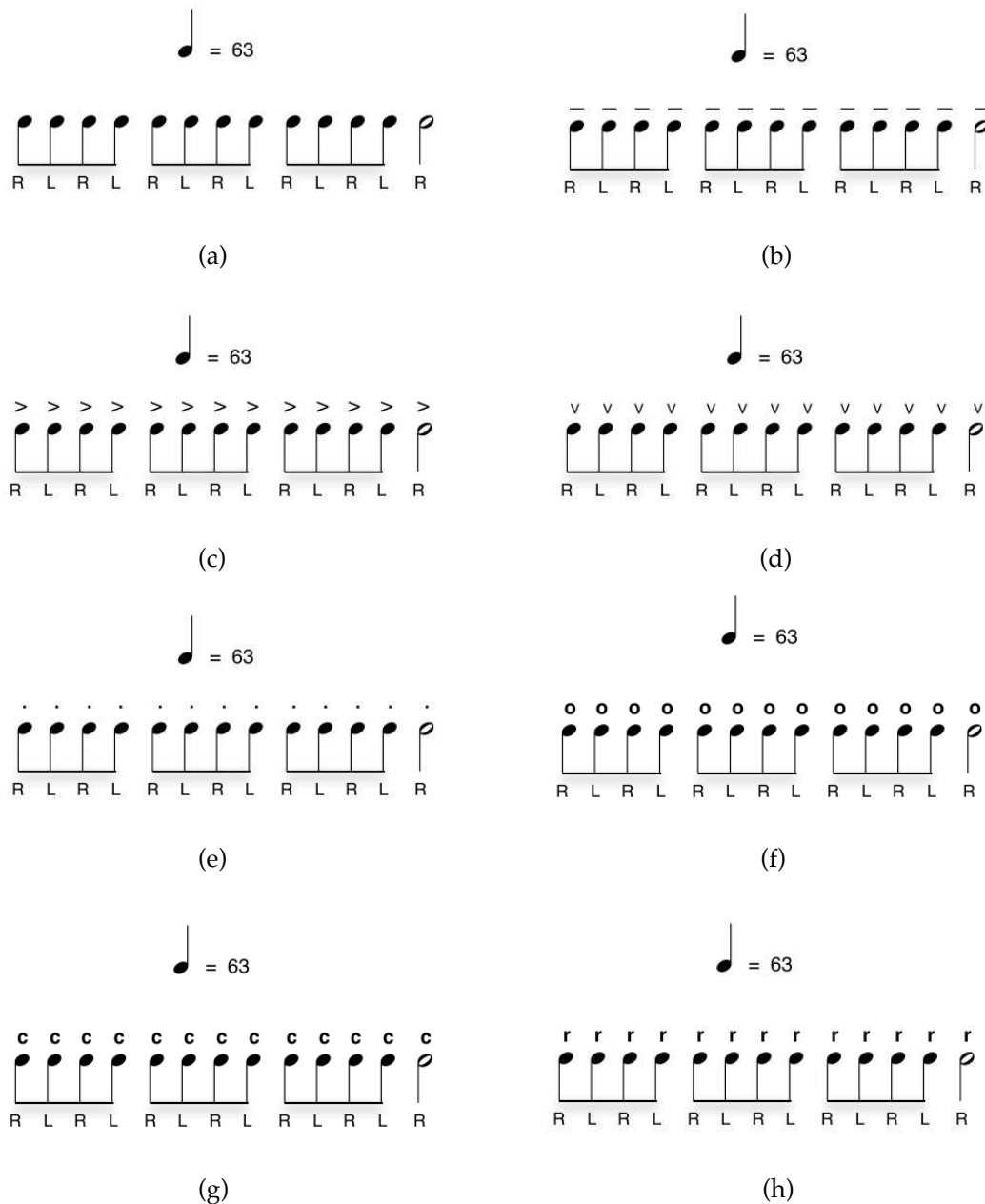


Figure 4: Validation exercises. (a) Sequence of legato strokes. (b) Sequence of tenuto strokes played at the rim. (c) Sequence of accent strokes. (d) Sequence of vertical accent strokes. (e) Sequence of staccato strokes. (f) Sequence of legato strokes played at the one-third. (g) Sequence of legato strokes played at the center. (h) Sequence of legato strokes played at the rim.

Extrapolation Exercises

The second category, extrapolation exercises, consists in musical excerpts that go beyond those obtained with motion capture. Such exercises were not performed by the musicians and therefore no corresponding articulation data are available. They typically mix various types of gestures and impact positions in one excerpt, as well as variations on the tempo of the performances, as depicted on Figure 5. We typically simulated several sequences of different gesture articulations. The major difference between exercises presented in Figure 5(a) and Figure 5(b) is that the former aims at testing the smooth sequencing of the gesture articulations, whereas the latter aims at testing the velocity and fidelity of quick changes between three particular articulations. We also simulated a well-known exercise consisting of an *accelerando-decelerando* of legato strokes - Figure 5(c) and Figure 5(d). This exercise aims at testing the ratio between the maximum velocity and the gesture fidelity our system can cope with. A final exercise – Figure 5(e) – consists in testing the sequencing of different beat impact locations for a given gesture articulation.

Qualitative Evaluation

A first musical evaluation was achieved for determining the overall degree of expertise of the virtual performer. From the resulting simulations, we can see that the virtual timpanist's performance is similar to that of a beginner/intermediate performer. Exercises have been performed with a wide and accurate collection of variations in impact locations and in the motion of the arm and forearm, even if only parts of the articulations were performed correctly.

Before going more deeply into the evaluation of the validation and extrapolation exercises, three general issues have been identified. A first issue, the capture limitation, is related to the data collected during our motion capture sessions. The capture limitation involves the lack of available data for the fingers

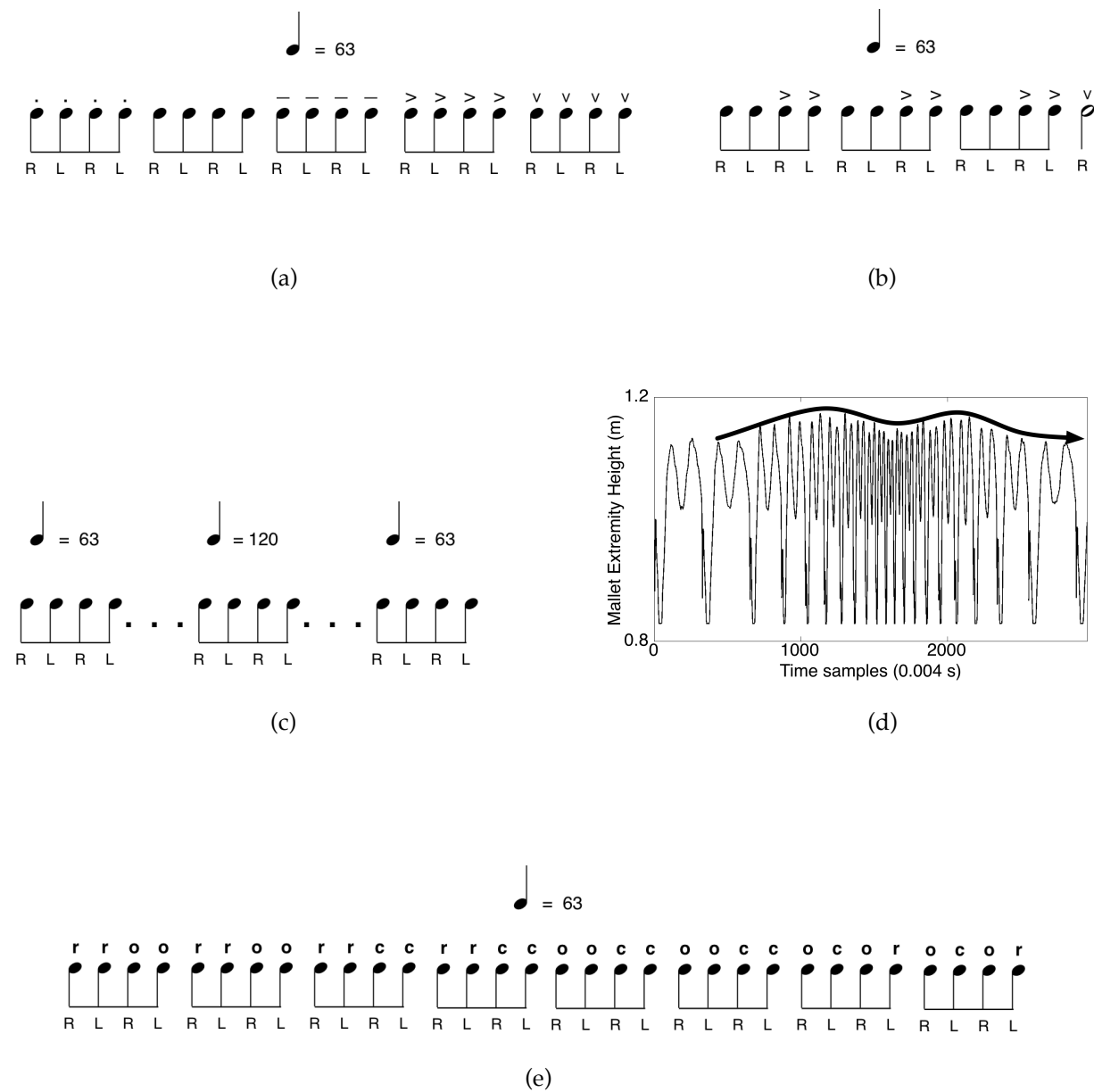


Figure 5: Extrapolation exercises. (a) Five gesture units repeated four times: staccato, legato, tenuto, accent and vertical accent. (b) Variations between legato and accent strokes, ending with a vertical accent. (c) Accelerando-decelerando of legato strokes. Note that the timpani database only has samples at a given tempo. (d) Height of the mallet extremity, the top arrow shows an amplitude increase, as well as a small drop when approaching the maximum accelerando velocity. (e) Sequence of legato strokes played at various locations on the drum membrane.

which are key components for the characterization of the grips. This can be an issue for highly-trained timpanists for evaluating how accurate the grip simulation is. The second issue, the gesture simulation limitation, is related to our physics-based simulation framework which apparently may cause exaggeration in upper-body movements, or conversely fixed movements for lower-body parts. It can then be difficult to discriminate between articulation modes, or to judge the body involvement of the virtual performer. Finally, a third issue, the sound limitation, is related to the Modalys membrane model that may not show as much variability as real timpani instrument. This can be difficult for instance for discriminating between some of impact locations. These three limitations can be found at different levels both for validation and extrapolation exercises.

Regarding validation exercises, the resulting simulations proved to show significant variations in the quality of the gesture articulations – exercises depicted by Figure 4(a-e). A much more satisfying comment is about impact locations simulations – exercises depicted by Figure 4(g-h) – where impact points were judged as accurate. However, the gesture simulation limitation has also been identified by an exaggeration of shoulders motion. A particular attention has also been paid on the sound outcome which was judged plausible and differential for the articulation modes, although the sound simulation limitation was pointed out as a factor of drifts in attack saturation and resonance compared to real timpani sounds.

As for extrapolation exercises, a first source of satisfaction is the comments obtained for exercises depicted by Figure 5(a-b). The transition between articulation modes (legato, tenuto, accent, vertical accent and staccato) has been judged as fairly smooth and realistic, with changes perceptible throughout the simulation. Discrepancies have however been noticed particularly for legato-tenuto and tenuto-accent pairs, so that visual differences between these are sometimes not very clear. This can be attributed to a combination of capture and gesture simulation

limitations, as unavailable finger effects as well as motion exaggeration may have altered part of the fine subtleties between these gesture modes. More interestingly about the *accelerando-decelerando* exercise, we can identify an height amplitude decrease of the mallet motion for respecting the beat velocity, as attested by the middle portion of the arrow in Figure 5(d). As for the quality of the gesture-sound interaction, a *crescendo* occurs naturally during the *accelerando* by setting the timpani membrane in resonance. It should be noticed that these features are also used by real timpanists to cope with roll speed and build-up a *crescendo* during a roll. This fact also attests to the natural response of the Modalys model of the drum membrane in response to the simulated legato strokes under an *accelerando* tempo variation. It shows moreover that the interaction model, and especially the collision detection of beat impacts, offers realistic features such as contact durations and forces in comparison to real timpani performances. This exercise also test the limits of our system, the speed/quality ratio of our system is somewhat reached with this exercise, since an irregular motion flow appears when approaching the maximum beat velocity, consequently the impact locations substantially change, causing inconsistency in the quality of the sounds produced. Finally, the exercise testing the sequencing of beat impact locations – Figure 5(e) – gave interesting results as differences both in location and sound output were judged as perceptible, and more specifically with respect to the end of the exercise, *i.e.* the two last groups of four strokes when switching between one-third, center, one-third and rim locations. However, it has been noted that center locations should have been more obvious among the three types of location, specifically for the centered strokes. The major reason for this effect is that the original recorded performances that were used to build and simulate this exercise were not performed accurately enough with respect to impact position. As we will discuss in the next section, the recorded beat impact locations - hollow symbols on Figure 6 - for the center location are too near to those

corresponding to the one-third location. This illustrates the important influence of the captured performance on our system, which will conserve performer's style but also drifts if any.

Discussion

Simulating instrumental percussion gestures for controlling sound synthesis processes presents advantages and limitations, and these are of different orders when considering each module of our framework. In the following sections we will analyze in detail some of the sources of variability that are produced by our simulation system and may be at the origin of unwanted artifacts, as well as new interesting possibilities. We therefore consider sequentially the advantages and limitations of, firstly, the instrumental gesture simulation and secondly the gesture editing step (*cf.* Figure 1).

Virtual Performer Simulation

We here underline the importance of the parameterization of the underlying physics model of virtual character. A non-optimal parameterization can lead to artifacts in the resulting synthesis of virtual percussionist gestures, as well as to effects on synthesized sounds. We also highlight the interest of simulating instrumental gestures with respect to the possible modulation of synthesized gestures while preserving the initial style and expressive characteristics of real percussion performances.

Advantages

From a simulation point of view, once an acceptable parameterization is achieved, it has been shown that associating a motion capture database and the physics-based synthesis of instrumental gestures yields accurate simulations (Bouënard 2009c,

Bouënard et 2010b).

A first advantage of the physics-based framework for simulating percussion gestures is the retrieval of the physical aspects (beat impacts) of the recorded instrumental gestures. No information was initially been recorded concerning beat impact durations and forces, our system therefore starts from pure kinematic data (joints position) to re-build a whole performance with physical interactions. A second advantage of our system is its virtual nature. Indeed, the proposed framework allows the modeling of realistic phenomena occurring during real percussion performances. Furthermore, it also allows going beyond reality, and provides the possibility of modifying the parameters of the virtual environment as wanted for creating new percussion performances. It is possible to explore changes in the physical model of the virtual percussionist, as well as in the interaction schemes between synthesized gestures and sound synthesis inputs through the contact information provided by the physical framework.

From a more musical point of view, an issue to take into account is the small variations in movements that are important to produce expressive performances. Indeed, deadpan virtual performances where movements are kept as close to an ideal movement target as possible are generally considered mechanical and non-expressive. Although our focus is more on the technical capabilities of our virtual performer, the smooth and non-mechanical interpretation of the simulated exercises is important. By choosing a single gesture unit for representing each playing mode in the various simulations (see section "Editing and Composition of Gesture Scores"), we might actually expect to eliminate the natural variability of movements from the original performance by the (real) percussionists. Nevertheless, thanks to the physics simulation layer, starting from a unique gesture unit for each gesture type will produce synthesized gestures with inherent variability due to the dynamic characteristic of the virtual character. Exploiting (while controlling) such drifts for

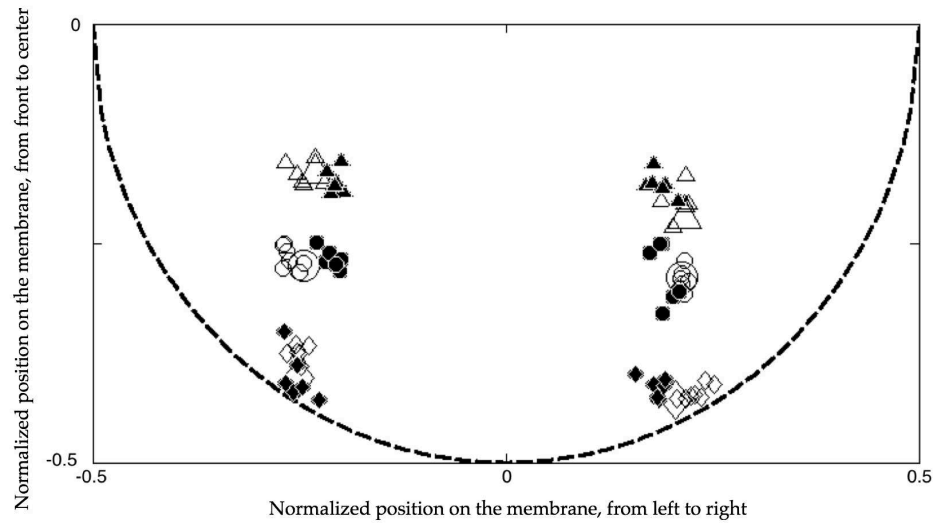


Figure 6: Comparison of impact locations: small hollow, large hollow and small bold symbols represent resp. strokes from the database, chosen strokes in the database and simulated impact locations.

slightly modifying the instrumental gesture may produce a form of expressiveness. The propagation (not the amplification) of such numerical drifts is also influenced by the editing and the articulation of gesture units. These small differences in gestural trajectories therefore lead to variations in the impact locations on the membrane, thus producing variations in corresponding synthesized sound. These different effects may be explored musically, and actually lead to non-mechanical performances by the virtual percussionist. An example is presented in Figure 6, comparing recorded and simulated legato strokes for the French grip. One can note the comparable variability of the (real) percussionist's strokes from motion capture (small hollow symbols), and the resulting variability of the synthesized strokes (small bold symbols) from multiple simulations of each gesture unit.

Limitations

Depending on the sequence and speed of the selected gestures, as well as to the fine

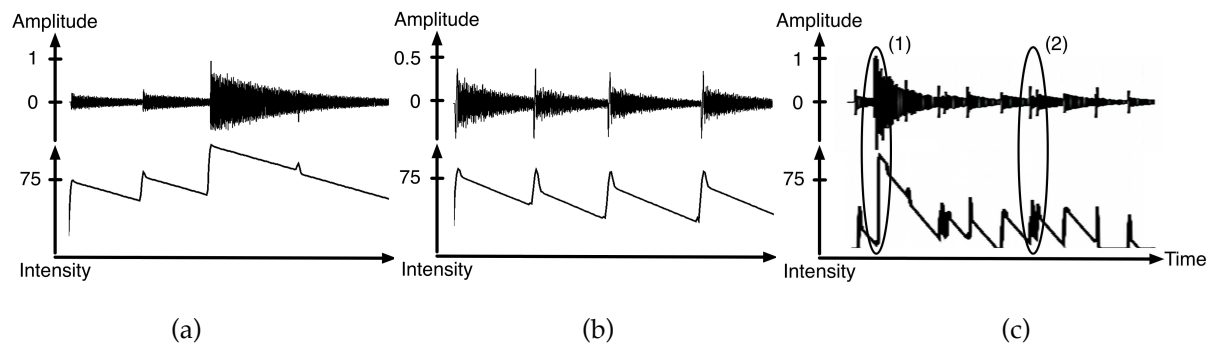


Figure 7: Sound waveform amplitude and intensity, from (a) a simulation artifact and (b) a stable simulation for the French grip. (c) As discussed in the text, a possible artifact for the German grip is when two close impacts are “glued” together by the sound synthesis, oval (1), and when impacts are finally heard as two, oval (2).

tuning of the mechanical joints, simulation artifacts can appear and result in large variations in beat impact positions and forces. These variations are due to the adaptation of the physics simulation to the constraints in the movement data, and may yield unexpected sound phenomena. An example is presented in Figure 7, where the virtual musician performs a sequence of four beats played legato with different sets of physical constraints. The parameterization of the mass and density of the different segments composing the physical model of the virtual percussionist should be chosen carefully, so that the inertial behavior of the system is comparable to that of humans. Otherwise unexpected artifacts may be obtained, *cf.* Figure 7(a) and Figure 7(b).

Moreover, it has been observed that percussion grips may have an influence on the quality of the simulations. While French grip-related simulations work well, German ones are less satisfactory. This can be explained by the fact that the German grip involves generally a technique requiring that the arms stay relatively immobile during a large portion of the gesture unit. This ends with a very stiff configuration of

the physics model for achieving this posture control that may lead to numerical instabilities. Moreover, the German grip requires more control from the wrist and fingers, features that are not taken into account in our model. Figure 7(c) shows the effect of this issue on the resulting sound. The virtual performer is given a task to perform a sequence of accent strokes using captured data of the German grip. From the resulting sounds, one can notice that at some moments, (1) the sound amplitude is larger than average, and (2) the resulting beats are not regularly spaced, although no dynamics or tempo variation was present in the score. Issue (1) characterizes a larger stroke as an artifact from the sound synthesis generation, *i.e.* the simulated gesture presents strokes that are very close temporally and spatially. The collision detection events, performed in a short lapse time, are then interpreted by the Modalys system which "glues" them together as a single yet much stronger stroke. As for issue (2), it can be explained by the limitations of both the capture database and the simulation framework, as discussed above.

Motion Database and Gesture Editing

Our framework depends highly on the motion clips initially recorded during the acquisition process of motion data. The editing processes associated with the motion database are off-line components that are of particular importance when considering the composition of a gesture score. We initially highlight the advantages of our framework in considering the simulation of percussion performances at the gestural level. We then discuss the completeness of the available motion capture database, as well as the editing method presented in this paper.

Advantages

In this work, we adopted a sketching approach to the composition of gesture scores made of elementary gesture units. Beginning and ending points of a gesture unit are

determined from mallet height trajectories. This constitutes an intuitive process where the user only needs to select and assemble a limited amount of canonical gesture units, one for representing each timpani playing variation. The whole gesture score can then be formed from copy-and-paste and concatenation operations on these gesture units.

Although the presented model does not take into account the subtle articulation mechanisms occurring in real percussion performances, a simple articulation of gesture units is implicitly achieved by the internal physics simulation when the gesture score simulation switches from a gesture unit to another. Our simple yet effective sketching process for composing gesture scores solely based on mallet extremity trajectories has the advantage of offering the articulation mechanism transparently to the user.

Another advantage of our framework in considering the composition at the gestural level is the preservation of the expressive features inherent to real percussion performances. As the composition process of gesture scores relies highly on real motion data, this insures that simulations will preserve the expressive components of the gesture units used in the gesture score. It allows consequently to preserve the style of the real performer in the simulated percussion performances. Extrapolation exercises can then be considered as supplementary exercises that we could have asked the percussionist to perform.

The presented motion capture database in section "Timpani Performance" focuses on a particular excerpt among the multitude of percussion techniques used by timpani performers. Our system provides the interesting possibility of enriching this limited database by the editing and simulation of novel percussion performances. One can therefore consider our approach as a mean of preventing the tedious and time-consuming task of capturing an exhaustive motion database containing all the possible combinations of playing techniques.

Limitations

The motion capture database presented in section "Timpani Performance" focuses on a specific set of timpani playing techniques, *i.e.* it provides gesture examples only for these playing techniques. As seen before, this fact limits the possible simulations to be performed with the proposed system. It also raises the question of the quality of the original captured motion. For instance we can see in Figure 6 the distribution of beat impacts from captured data is not as accurate as expected, with center locations too near from one-third locations. It leads to simulations showing the same trend, as pointed out in the qualitative evaluation.

Another limiting issue is the choice of the set of gesture units used during the editing step (section "Editing and Composition of Gesture Scores"). Such issue yields inevitably to the question of the existence of a best-suited gesture unit as well as its automatic recognition, compared to other examples of the same playing variation. If such "best" gesture unit exists, how can one identify it among the numerous replicates available in the motion database? Otherwise, does it make sense to average multiple gesture units so that a "mean" gesture unit can be used? We believe that using an average gesture might lead to the loss of some of the expressive features implicitly contained in the gesture signal. We therefore chose one specific gesture example within the sequence, without trying to find an optimized criteria for automatically identifying and selecting it.

A final issue related to the editing process is the concatenation of gesture units, which is the basis of the composition process. The gesture composition presented in this paper reveals only a sketching approach for composing gesture scores. It could be improved by considering higher-level planning processes responsible for the sequencing of gestures. In real percussion playing, high-level control mechanisms are indeed in charge of the articulation between playing modes, depending on many factors such as playing style, tempo variations as well as

expressive nuances.

Conclusion and Future Work

We presented a system for synthesizing percussion performances, from the editing process of gesture scores, to the physics simulation of a virtual percussionist that controls sound synthesis processes. The approach was applied to the simulation of both validation and extrapolation exercises consisting in pre-recorded as well as new edited percussion sequences. The qualitative evaluation of both validation and extrapolation exercises by a university percussion professor have attested the overall realism of the resulting simulations. Our system is capable of handling realistic motion strategies, such as performing gesture articulations separately, as well as their smooth sequencing. Our system can also cope with motion-sound strategies that are tightly linked. The best example is the *accelerando-decelerando* exercise which showed that the virtual performer could adapt to high tempi and adequately decrease his mallet height while performing a sound crescendo, which are real-world properties of a percussion roll. timing properties. The evaluation also pointed out three major issues that make room for improvements for our gesture-sound synthesis system.

The capture limitation yielded to difficulties to recognize some gesture articulations in our simulated exercises. Although our motion capture setup does capture the orientation of hands and mallets, such information appears to be unsufficient for modeling accurately the grasp subtleties that could enhance the realism of the resulting grip simulations. This issue could be overcome by adding sensors on performer's fingers, and therefore monitoring the grasp mechanisms involving the fingers and the mallets. Of course we can consequently anticipate an additional work for extending our virtual performer modeling and simulation to take into account such additional data.

An other issue is related to the gesture simulation limitation, leading to upper-body motion exaggeration, as well as fixed movements for lower-body parts. Motion exaggeration can be explained by the dynamic properties of the body parts of the virtual percussionist. When the physical model is put into motion, these body parts seem to acquire too much inertia, so that even if the overall simulated motion is accurate enough compared to the recorded gesture units (Bouënard 2009c, Bouënard et al 2010b), an exaggeration of motion appears. An exploration of the dynamic properties of the physics model could solve this issue, more specifically joints linking limbs could involve an increase in stiffness. Moreover, fixed movements of the lower-body are related to balance strategies, *i.e.* how a timpanist involves his center of mass for helping him to switch between playing modes. In fact no balance strategies are involved in the simulation of the percussion exercises, although information on performers' center of mass was captured in the motion capture sessions. This simulation choice explains the remark on the quite static bottom part of the virtual performer. Balance strategies could also help us to extend the repertoire of the virtual performer, as they are important if we want to simulate exercises where the virtual performer plays on several timpani instruments.

A final issue is the sound simulation limitation, which sometimes leads to resulting sounds without the metallic grain of real timpani sounds, or with too much attack saturation and not enough resonance. The lack of sound resonance is easily explained by the fact that the drum membrane model does not feature a resonance body. Moreover, the non-metallic nature of the resulting sounds may also come from a combination of imprecisions both from the drum membrane model itself since no metallic rim is included in the Modalys model, and from beat impact locations inaccuracies. A solution would then involve considering an extended model of the timpani instrument with Modalys.

The overall musical perspectives of our gesture-sound synthesis system are

encouraging as we plan also to apply it both for musical pedagogy and composition. A pedagogical tool based on our system would offer a multimodal interface for studying timpani gestures of a percussionist, being a student or a professor. Of course it would first imply a minimal amount of motion capture sessions, which can involve a heavy and time-consuming setup, but we believe it is a price to pay for accessing such pedagogical capabilities. A compositional tool can also be derived, with which composers could then choose for a substantial palette of percussion variations, such as gesture articulations, beat locations and tempi, yet from a minimal amount of gesture units. Gesture score inputs could be graphically implemented in a user interface for easing the editing process of composers, while giving them both visual and audio feedback of their simulated composition.

References

- Adrien, J.-M. 1989. "Etudes de Structures Complexes Vibrantes, Application à la Synthèse par Modèles Physiques". *PhD thesis, Université Paris VI, France.*
- Avanzini, F. 2007. "Synthesis of Environmental Sounds in Interactive Multimodal Systems". *Proc. of the Int. Conference on Auditory Display*, pp. 181-188.
- Avanzini, F., and D. Rochesso. 2004. "Physical Modeling of Impacts: Theory and Experiments on Contact Time and Spectral Centroid". *Proc. of the Int. Conference on Sound and Music Computing*, pp. 287-293.
- Benning, M. S., A. Kapur, B. Till, G. Tzanetakis and P. Driessen. 2007. "A Comparative Study on Wearable Sensors for Signal Processing on the North Indian Tabla". *Proc of the IEEE Conf. on Communications, Computers and Signal Processing*, pp. 621-624.
- Bouënard, A., S. Gibet, and M. M. Wanderley. 2008a. "Enhancing the Visualization of Percussion Gestures by Virtual Character Animation". *Proc. of the Int. Conference on New Interfaces and Musical Expression*, pp. 38-43.

- Bouënard, A., M. M. Wanderley, and S. Gibet. 2008b. "Analysis of Percussion Grip for Physically-based Character Animation". *Proc. of the Int. Conference on Enactive Interfaces*, pp. 22-27.
- Bouënard, A., S. Gibet, and M. M. Wanderley. 2009a. "Hybrid Motion Control combining Inverse Kinematics and Inverse Dynamics Controllers for Simulating Percussion Gestures". *Proc. of the Int. Conference on Computer Animation and Social Agents*, pp. 17-20.
- Bouënard, A., M. M. Wanderley, and S. Gibet. 2009b. "Advantages and Limitations of Simulating Percussion Gestures for Sound Synthesis". *Proc. of the Int. Computer Music Conference*, pp. 255-261.
- Bouënard, A. 2009c. "Synthesis of Music Performances: Virtual Character Animation as a Controller of Sound Synthesis". *PhD thesis, European University of Brittany – University of South Brittany, France*.
- Bouënard, A., M. M. Wanderley, and S. Gibet. 2010a. "Gesture Control of Sound Synthesis: Analysis and Classification of Percussion Gestures". *Acta Acustica united with Acustica*, 96(4):668-677.
- Bouënard, A., S. Gibet and M. M. Wanderley. 2010b. "Hybrid Inverse Motion Control for Virtual Characters Interacting with Sound Synthesis – Application to Percussion Motion". *The Visual Computer Journal*, *accepted*.
- Bregler, C. 1997. Learning and Recognizing human Dynamics in Video Sequences. *Proc of the Int. Conference on Computer Vision and Pattern Recognition*, pp. 568-575.
- Carter, E. 1968. "Eight Pieces for Four Timpani (One Player)". *Associated Music Publishers, New York (USA)*.
- Chuchacz, K., S. O'Modhrain, and R. Woods. 2007. "Physical Models and Musical Controllers: Designing a Novel Electronic Percussion Instrument". *Proc. of the Int. Conference on New Interfaces for Musical Expression*, pp. 37-40.

- Dahl, S., 2004. "Playing the Accent: Comparing Striking Velocity and Timing in Ostinato Rhythm Performed by Four Drummers". *Acta Acustica united with Acustica*, 90(4):762-776.
- Dobrian, C., and D. Koppelman. 2006. "The 'E' in NIME: Musical Expression with New Computer Interfaces". *Proc. of the Int. Conference on New Interfaces for Musical Expression*, pp. 277-282.
- Eigenfeldt, A., and A. Kapur. 2008. "An Agent-based System for Robotic Musical Performance". *Proc. of the Int. Conference on New Interfaces for Musical Expression*, pp. 277-282.
- Ellis, N., J. Bensoam, and R. Caussé. 2005. "Modalys Demonstration". *Proc. of the Int. Computer Music Conference*, pp. 101-102.
- Friese, A., and A. Lepak. 1985. "The Complete Timpani Method". *Alfred Music Publishing, Van Nuys (USA)*.
- Gibet, S., and J.-L. Florens. 1988. "Instrumental Gesture Modeling by Identification with Time-varying Mechanical Models". *Proc. of the Int. Computer Music Conference*, pp. 28-40.
- Gibet, S., and P.-F. Marteau. 1990. "Gestural Control of Sound Synthesis". *Proc. of the Int. Computer Music Conference*, pp. 387-391.
- Green, R.D., and L. Guan. 2004. "Quantifying and Recognizing Human Movement Patterns from Monocular Video Images: Part I". *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):179-190.
- Hänninen R., L. Savioja, and T. Takala. 1996. "Virtual Concert Performance – Synthetic Animated Musicians playing in an Acoustically Simulated Room". *Proc. of the Int. Computer Music Conference*, pp. 402-404.
- Hoffman, G., and G. Weinberg. 2010. "Shimon: An Interactive Improvisational Robotic Marimba Player". *Proc. of the Int. Computer-Human Interaction Conference*, pp. 3097-3102.

- Instrumentalist. 1984. "Percussion Anthology: A compendium of Articles from the Instrumentalist". *Instrumentalist Co (3rd Edition), Evenston (USA)*.
- Kapur, A., G. Essl, P. Davidson, and P. Cook. 2003. "The Electronic Tabla Controller". *Journal of New Music Research*, 32(4):351-359.
- Kapur, A., M. Darling, M. Wiley, O. Vallis, J. Hochenbaum, J. Murphy, D. Diakopoulos, C. Burgin, and T. Yamin. 2010. "The Machine Orchestra". *Proc of the Int. Computer Music Conference*, pp. x-x.
- Lytle, W. 2001. "Pipe Dream". *SIGGRAPH Computer Animation Festival*.
- Mäki-Patola, T. 2005. "User Interface Comparison for Virtual Drums". *Proc. of the Int. Conference on New Interfaces for Musical Expression*, pp. 144-147.
- Miranda, E., and M. M. Wanderley. 2006. "New Digital Music Instruments: Control and Interaction beyond the Keyboard". *A-R Editions*.
- O'Brien, J., C. Shen, and C. Gatchalian. 2002. "Synthesizing Sounds from Rigid Body Simulations". *Proc. of the Symposium on Computer Animation*, pp. 175-181.
- Ramstein, C. 1991. "Analyse, Représentation et Traitement du Geste Instrumental". *PhD thesis, Institut National Polytechnique de Grenoble, France*.
- Tindale, A., A. Kapur, G. Tzanetakis, P. Driessen, and A. Schloss. 2005. "A Comparison of Sensor Strategies for Capturing Percussive Gestures". *Proc. of the Int. Conference on New Interfaces for Musical Expression*, pp. 200-203.
- van den Doel, K., P. Kry, and D. Pai. 2001. "FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Simulation". *Proc. of the Annual Conference on Computer Graphics and Interactive Techniques*, pp. 537-544.
- Wagner, A. 2006. "Analysis of Drumbeats – Interaction between Drummer, Drumstick and Instrument". *Master thesis, KTH Royal Institute of Technology, Sweden*.
- Weinberg, G., and S. Driscoll. 2006. "Toward Robotic Musicianship". *Computer Music Journal*, 30(4):28-45.